

Surrogacy of progression free survival for overall survival in metastatic breast cancer studies: meta-analyses of published studies

Madan G. Kundu and Suddhasatta Acharyya

Novartis Pharmaceutical Corporation
East Hanover, NJ, USA

Abstract

Purpose: PFS is often used as a surrogate endpoint for OS in metastatic breast cancer studies. We have evaluated the association of treatment effect on PFS with significant HR_{OS} (and how this association is affected by other factors) in published prospective metastatic breast cancer studies.

Methods: A systematic literature search in PubMed identified prospective metastatic breast cancer studies. Treatment effects on PFS were determined using hazard ratio (HR_{PFS}), increase in median PFS (ΔMED_{PFS}) and % increase in median PFS ($\% \Delta MED_{PFS}$). Diagnostic accuracy of PFS measures (HR_{PFS} , ΔMED_{PFS} and $\% \Delta MED_{PFS}$) in predicting significant HR_{OS} was assessed using receiver operating characteristic (ROC) curves and classification tree approach (CART).

Results: Seventy-four cases (i.e., treatment to control comparisons) from 65 individual publications were identified for the analyses. Of these, 16 cases reported significant treatment effect on HR_{OS} at 5% level of significance. Median number of deaths reported in these cases were 153. Area under the ROC curve (AUC) for diagnostic measures as HR_{PFS} , ΔMED_{PFS} and $\% \Delta MED_{PFS}$ were 0.69, 0.70 and 0.75, respectively. Classification tree results identified $\% \Delta MED_{PFS}$ and number of deaths as diagnostic measure for significant HR_{OS} . Only 7.9% (3/39) cases with ΔMED_{PFS} shorter than 48.27% reported significant HR_{OS} . There were 7 cases with ΔMED_{PFS} of 48.27% or more and number of deaths reported as 227 or more – of these 5 cases reported significant HR_{OS} .

Conclusion: $\% \Delta MED_{PFS}$ was found to be a better diagnostic measure for predicting significant HR_{OS} . Our analysis results also suggest that consideration of total number of deaths may further improve its diagnostic performance. Based on our study results, the studies with 50% improvement in median PFS are more likely to produce significant HR_{OS} if the total number of OS events at the time of analysis is 227 or more.

Keywords: Metastatic breast cancer, Progression free survival, Overall survival, surrogacy, meta-analysis, ROC curve, classification tree, Surrogate threshold effect.

Introduction

As per national cancer institute, in the U.S., breast cancer is the second most common non-skin cancer and the second leading cause of cancer-related deaths in women; and, therefore, there has always been a high demand for novel breast cancer therapies. At the time of preparing this manuscript, based on ClinicalTrial.gov search [1], 175 phase III breast cancer studies were actively recruiting patients. For breast cancer therapies, the main goal is to improve overall survival (OS) and quality of life [2, 3]. US FDA guideline [4] states that “[overall] survival is considered the most reliable cancer

endpoint”. Due to the advancement in metastatic breast cancer management and therapies, there has been marked improvement in OS in breast cancer patients in the last few decades. Consequently, patients need to be followed-up for longer period of time to observe sufficient number of OS events (i.e., deaths) [5] before treatment effect on OS can be evaluated statistically. Further, as many patients switch to second line (and beyond) therapies upon progression, the OS time may be influenced by post-progression therapy. For these reasons, surrogate endpoints such as progression-free survival (PFS) or objective response rate (ORR) are being increasingly used for accelerated approvals, with PFS being the one used most often [2]. The basis for using PFS as surrogate endpoint for OS is as follows: cancer progression represents an ominous march toward death from malignancy. Hence, the longer it takes for the cancer to progress, the longer a patient will live [6]. In general, PFS has not been statistically validated for surrogacy of OS yet in breast cancer studies[2, 4]. Reported results regarding association between Hazard ratio of PFS and OS in the metastatic breast cancer studies are mixed: For example, Hackshaw *et al.* [7] found a correlation of 0.87; Burzykowski *et al.* [8] reported correlation of 0.48; Michiels *et al.* [9] reported R^2 (i.e. proportion of the variance in the true endpoint that is explained by the surrogate endpoint) as 0.51.

According to Prentice’s definition[10], in order for PFS to be a “statistically validated” surrogate endpoint for OS, “test for null hypothesis of no treatment effect in PFS” should be a valid “test for null hypothesis of no treatment effect in OS”. The test for treatment effect on OS is carried out by testing $HR_{OS}=1$, where HR_{OS} is the hazard ratio (HR) of OS. However, many randomized clinical trials failed to demonstrate significant treatment effect in OS despite demonstrating significant treatment effect in PFS. The current project attempts to investigate the trial level surrogacy in breast cancer studies from a diagnostic testing perspective using nonparametric approaches. It is important to note that our investigation differs from previous investigations [7, 8, 11, 12, 13, 14, 15, 16] based on meta-analytic methods, where the primary purpose was to examine the strength of treatment effect on PFS to predict treatment effect on OS at trial level. The definition of trial level surrogacy in the current investigation is intuitive and aligned with the ultimate question that all stakeholders, regulators in particular, are often seeking an answer to, from a phase III cancer clinical trial – Is there a statistically significant OS benefit in the new treatment that is discernible from the data on progression-free survival (PFS) in metastatic breast cancer studies? This definition of trial level surrogacy was also considered by Burzykowski and Buyse [17] as it can be useful to estimate the “Surrogate threshold effect”. Surrogate threshold effect can be defined as the minimum treatment effect on PFS measure that is required to predict statistically significant HR_{OS} .

Our goal was to evaluate the trial level surrogacy of PFS for OS solely based on published clinical trial results. Burzykowski *et al.* [8] evaluated trial level surrogacy by fitting simple (log-) linear regression analysis to model HR_{OS} with ratio of median PFS time and then used R^2 as a measure of trial level surrogacy. Buyse *et al.* [18] proposed to estimate trial level surrogacy using R^2 as well, but in a more sophisticated way using trial specific random effects. These methods make various model assumptions such as PFS and OS are linearly associated [8] or some distributional assumption [18]). As Venook and Tabernero [6] have pointed out association of PFS with OS may be complicated in today’s era and, therefore, a simplified linear model may not be sufficient to describe the association. Further, the use of R^2 is heavily impacted by the presence of outlier [19]. Another problem related to R^2 is the difficulty in interpreting its value [17]. For these reasons, we have adopted non-parametric approaches to evaluate the trial level surrogacy which, unlike parametric methods, do not require to make distributional assumptions or to pre-specify the form of the association. The advantages

Table 1: Summary of publications by journal and year

| Journal of Publication | 2014-2015 | 2012-2013 | 2010-2011 | 2005-2009 | 2000-2004 | Total |
|----------------------------------|-----------|-----------|-----------|-----------|-----------|-------|
| Journal of Clinical Oncology | 7 | 9 | 4 | 5 | 3 | 28 |
| Breast Cancer Research Treatment | 3 | 2 | 4 | 2 | | 11 |
| Annals of Oncology | 5 | 3 | | | 1 | 9 |
| Cancer | 1 | | 1 | 1 | | 3 |
| Clinical Breast Cancer | | | 3 | | | 3 |
| Clinical Cancer Research | | 2 | | | | 2 |
| Others | 1 | 4 | 1 | 2 | 1 | 9 |
| Total | 17 | 20 | 13 | 10 | 5 | 65 |

of non-parametric methods are that these methods are completely data-driven and free from model assumptions. Consequently, non-parametric methods have obvious advantage of producing results which are solely based on observed data and are not dependent on unverifiable model assumptions. Non-parametric methods can be also useful (a) to find out which PFS measure is relatively more important in predicting significant HR_{OS} , (b) to study the influence of other factors (e.g., sample size and total number of events) on the association of PFS measure and significant HR_{OS} as, for example, the power for statistical test of HR_{OS} is a function of total number of OS events, and (c) to estimate surrogate threshold effect. Results from non-parametric methods are often easy to interpret, and allow granular visualization of the results. For this project, breast cancer studies were our focus, but the similar investigation can be carried out for other indications as well. Throughout the article, (unless otherwise mentioned), ‘statistically significant’ would imply that the significance was in favor of the treatment.

Methods

Literature search

A systematic literature search in PubMed (July 2015) was performed to identify published prospective studies on metastatic breast cancer research with both PFS and OS comparison results reported. The search syntax used was as follows: “(((Breast Cancer[Title]) AND Randomized[Title/Abstract]) AND Progression free survival[Text Word]) AND Overall survival[Text Word]”. The PubMed search returned 181 publications between Jul-2000 and Jul-2015. Many of these studies were systematic literature review or meta-analyses and hence dropped. Further, studies with either PFS or OS not reported were also excluded. We were able to find 64 individual prospective studies [20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84] where both PFS and OS comparison results were reported. In addition, in one publication [74], instead of PFS, time to progression (TTP) was reported and that study was included. Therefore, we had total of 65 publications for the meta-analyses.

Data extraction

Of the 65 selected publications, in seven prospective studies [25, 26, 27, 57, 59, 61, 84], two pairs of treatment-to-control comparisons were reported and in one prospective study [35], three pairs of treatment-to-control comparisons were reported. Therefore, we had total of 74 treatment-to-control comparison available for the meta-analyses. For each treatment-to-control comparison, the following information were extracted: randomization status, blinding status (open or blinded), total sample size (treatment plus control), total number of events (treatment plus control), median PFS, median OS, HR (hazard ratio) in PFS (HR_{PFS}), HR in OS HR_{OS} , reported p-value (or significance status) for HR_{PFS} and reported p-value (or significance status) for HR_{OS} . In case both local and central PFS assessments were reported, the one which was reported as primary endpoint was considered.

Statistical methods

Treatment effect on PFS was determined using the following measures: hazard ratio (HR_{PFS}), increase in median PFS (ΔMED_{PFS}) and % increase in median PFS ($\% \Delta MED_{PFS}$). All three measures were used as diagnostic tools for predicting statistically significant HR_{OS} in favor of treatment (yes/no).

We have assessed the trial level surrogacy of PFS for OS by evaluating the diagnostic accuracy of these comparative PFS measures to predict statistically significant HR_{OS} . Diagnostic accuracy of comparative PFS measures (HR_{PFS} , ΔMED_{PFS} and $\% \Delta MED_{PFS}$) in predicting significant HR_{OS} was assessed using receiver-operating characteristic (ROC) curve [85], and classification tree (using CART algorithm [87]) approach. Empirical ROC curves were drawn plotting the true positive rate (proportion of correct prediction of significant HR_{OS} based on comparative PFS measure among those reporting significant treatment effect on HR_{OS}) against the false positive rate (proportion of wrong prediction of significant HR_{OS} based on PFS measure among those reported non-significant treatment effect on HR_{OS}). True positive rate, and false positive rate were obtained at each unique value of comparative PFS measures. For a given unique value of x , if comparative PFS measure was greater than or equal to x , then it was predicted that HR_{OS} will be significant; otherwise not. The accuracy of the diagnostic measure was assessed by numerically computing the area under ROC curve (AUC), with larger AUC implying better accuracy. Optimal cut-off points based on ROC curve were identified according to Youden's index [88]. According to Youden's criteria a optimum cut-off point for prediction of significant HR_{OS} would be one that maximizes the difference between true positive rate and false positive rate.

We have utilized classification tree to answer following questions: (a) which trial level measure of treatment benefit in PFS has stronger association with significant HR_{OS} in favor of treatment – HR_{PFS} or (%) median improvement in PFS? (b) Is there any other factor(s) (e.g., total number of deaths) that influence significance of HR_{OS} ? (c) if yes, then how does this measure modify the association of treatment benefit in PFS with significant HR_{OS} ? The following variables were used as partitioning variables in the classification tree analysis: all three comparative PFS measures (HR_{PFS} , ΔMED_{PFS} and $\% \Delta MED_{PFS}$), total sample size and total number of reported deaths. *Bagging* method [89] was applied to identify the most important partitioning variable(s). All statistical analyses were performed using R 3.0.2. A two-sided p value of $< .05$ was considered statistically significant. ROC analysis was carried out using "ROCR" package [90]. Classification tree was

Table 2: Summary of 74 comparisons (i.e. treatment to control comparisons) included in the meta-analyses

| Characteristics | |
|---|------------------------|
| Study phase – n(%) | |
| Phase III | 47 (63.5%) |
| Phase II/IIB | 19 (25.7%) |
| Unknown | 8 (10.8%) |
| Blinding status – n(%) | |
| Open | 38 (51.4%) |
| Blinded | 15 (20.3%) |
| Unknown | 21 (28.3%) |
| Sample size (n=74) | |
| Median (Min, Max) | 259 (41, 1349) |
| Number of deaths (n=60) | |
| Median (Min, Max) | 153 (19, 997) |
| Type of control – n(%) | |
| Active | 68 (91.9%) |
| Placebo | 3 (4.1%) |
| Standard care | 3 (4.1%) |
| Line of therapy – n(%) | |
| First line | 49 (66.2%) |
| 2nd or beyond | 25 (33.8%) |
| Increase in median PFS, ΔMED_{PFS} (n=72) | |
| Median (Min, Max) | 1.60 (-0.50, 10.90) |
| % increase in median PFS, $\%\Delta\text{MED}_{PFS}$ (n=72) | |
| Median (Min, Max) | 29.99 (-10.42, 294.60) |
| HR in PFS, HR_{PFS} (n=68) | |
| Median (Min, Max) | 0.78 (0.24, 1.18) |
| HR in OS, HR_{OS} (n=63) | |
| Median (Min, Max) | 0.85 (0.37, 1.49) |

constructed using “rpart” package [91] and for *bagging* method we have used “randomForest” package [92].

Results

Description

We had a total of 74 treatment-to-control comparisons available from 65 publications for the meta-analyses. The majority of these publications were published in the Journal of Clinical Oncology (28; 43%), Breast cancer research treatment (11; 17%) and Annals of Oncology (9; 14%), see Table 1. The majority (44, 68%) of these studies recruited patients to treat as first line therapy. Forty-one (63%) of these studies were phase III. In 60 studies comparison was made with active control, in 2 studies comparison was made with placebo and in remaining 3 studies standard care was used as comparator.

Table 3: Number of cases (i.e. treatment to control comparisons) reporting significant (at 5% level) difference in PFS and OS time

| Progression free survival (PFS) | Overall survival (OS) | |
|-----------------------------------|------------------------------|----------------------------------|
| | HR _{OS} significant | HR _{OS} not significant |
| HR _{PFS} significant | 12 | 21 |
| HR _{PFS} not significant | 4 | 37 |
| Level of significance is 5%. | | |

The characteristics of the 74 comparisons are summarized in Table 2. Of the 74 comparisons, 73 (98.6%) were reported to be made in randomized set-up and 47 (63.5%) were based on phase III trials. Only 15 (20.3%) comparisons were reportedly carried out in blinded fashion and blinding status was not reported for 21 (28.3%) comparisons. The median total sample size was 259 and the median number of deaths reported was 153. The majority (91.9%) of the comparisons included active control in the study, and in 66.2% comparisons, treatment under investigation was the first line therapy. The median HR_{PFS} and HR_{OS} were 0.78 and 0.85, respectively. Further, on average, median PFS time was increased by 1.60 months which translates to 29.99% increase in median PFS time.

Of the 74 comparisons, significant (at 5% level) HR_{PFS} and HR_{OS} were reported in 33 (44.6%) and 16 (21.6%) cases, respectively (see Table 3). The comparisons with significant HR_{PFS} are 5.29 times more likely to have significant HR_{OS} compared to the comparisons where HR_{PFS} was not reported as significant. However, more importantly, only 36.4% (12/33) of comparisons with significant HR_{PFS} also reported significant HR_{OS}.

Diagnostic accuracy (using ROC analysis)

ROC curves for each of HR_{PFS}, ΔMED_{PFS} and $\%\Delta\text{MED}_{PFS}$ evaluating diagnostic accuracy to predict significant HR_{OS} are displayed in Figure 1. AUC from ROC curves based on diagnostic measure of HR_{PFS} (AUC=0.69) and ΔMED_{PFS} (AUC=0.70) were numerically close. However, $\%\Delta\text{MED}_{PFS}$ offers relatively better diagnostic accuracy with AUC as 0.75. From the ROC curve of $\%\Delta\text{MED}_{PFS}$ in Figure 1, the optimal cut-off point (according to Youden’s index) was 44.83%, for which the sensitivity (i.e. true positive rate) was 81.3% and specificity (i.e. 1-false positive rate) 76.8%. It can be interpreted as follows: if we set a predictive rule to classify the cases with improvement in median PFS greater than 44.83% as producing significant HR_{OS} subsequently, then 81.3% of cases reporting significant HR_{OS} will be correctly predicted and 76.8% of cases reporting non-significant HR_{OS} will be correctly predicted. Another cut-off point of interest could be 33.33% for which the sensitivity and specificity were 87.5% and 62.5%, respectively.

Diagnostic accuracy (using classification tree)

In classification tree approach, in addition to comparative PFS measures, number of deaths and sample size were also considered as predictor variables. Classification tree results based on 74 cases (i.e.

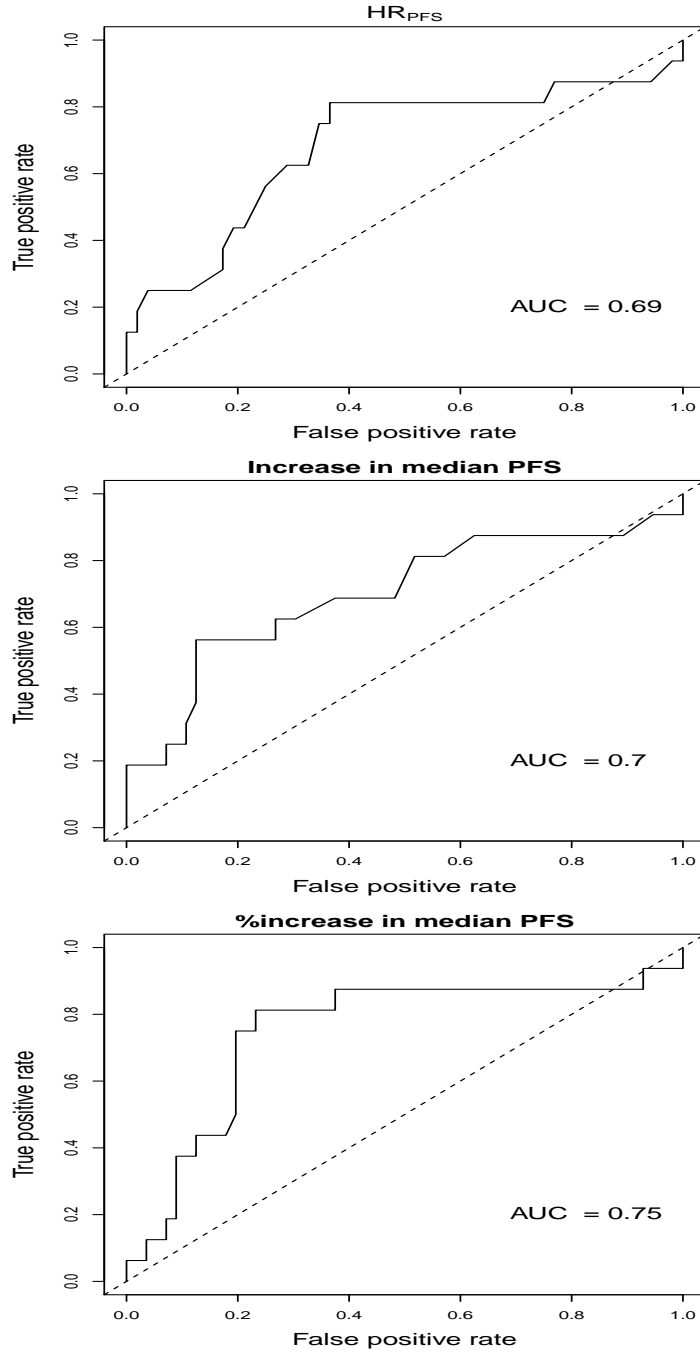


Figure 1: ROC curves using treatment effect on PFS as diagnostic measure for prediction of significant HR_{OS} at 5% level of significance. Treatment effect on PFS was assessed using (a) hazard ratio (HR_{PFS}), (b) increase in median PFS (ΔMED_{PFS}) and (c) % increase in median PFS ($\% \Delta MED_{PFS}$). True positive rate was defined as proportion of correct prediction among the comparisons reporting significant HR_{OS} . False positive rate was defined as (% of wrong prediction among the comparisons reporting non- significant HR_{OS}).

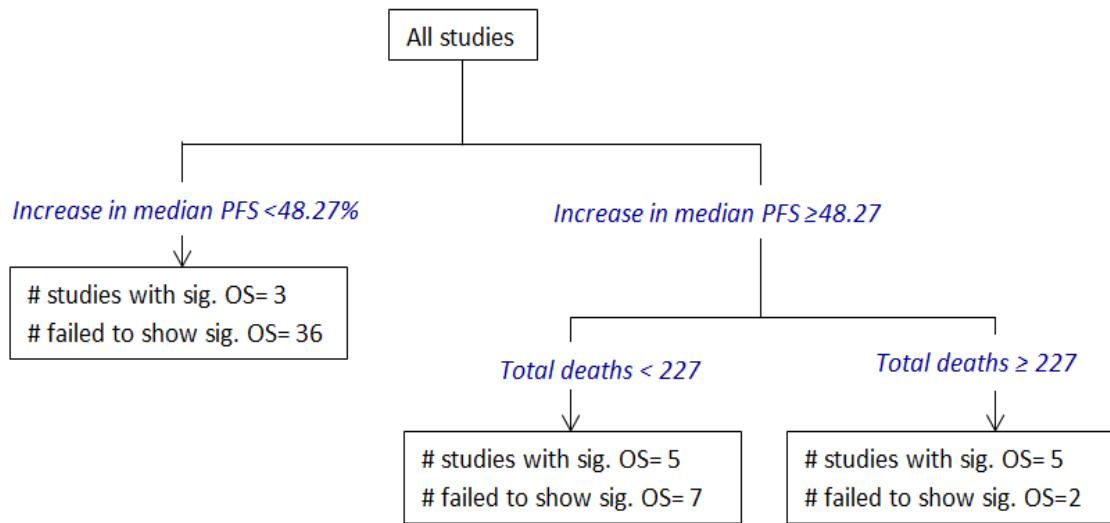


Figure 2: Classification tree for predicting significant treatment effect on OS

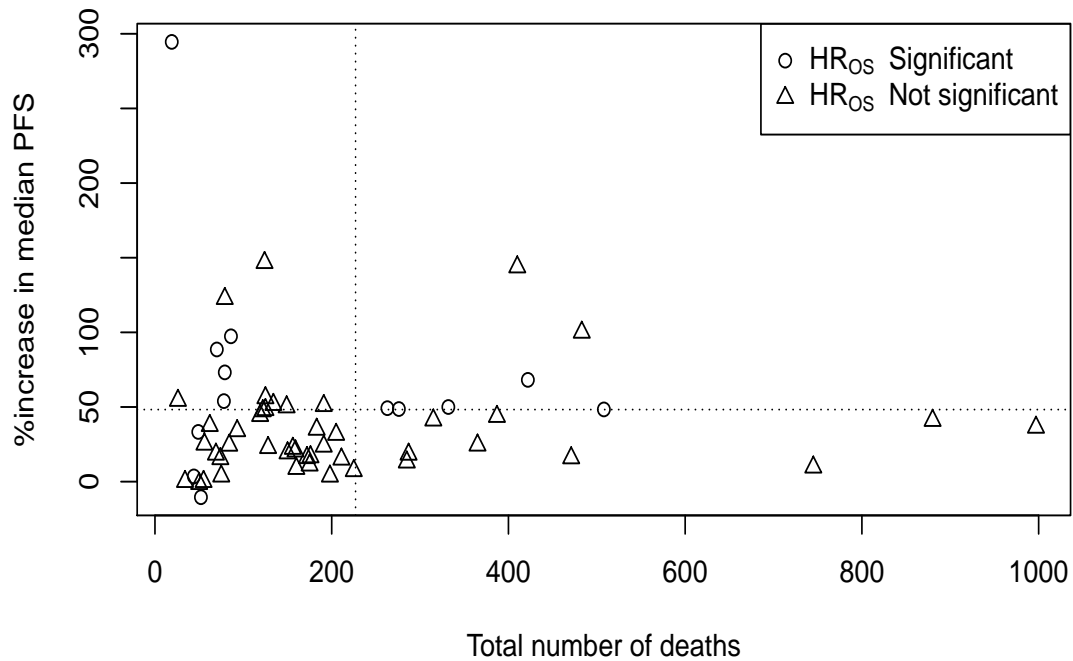


Figure 3: Display of comparisons with statistical significance status of HR_{OS} in terms of % increase in median PFS and total number of deaths

comparisons) identified $\% \Delta \text{MED}_{PFS}$ and total number of deaths as diagnostic measures for significant HR_{OS} . Importantly, we found $\% \Delta \text{MED}_{PFS}$ as a more important predictor variable compared to HR_{PFS} or ΔMED_{PFS} . This is very much consistent with our findings observed in the analysis based on the ROC curve. The *bagging* method also suggested $\% \Delta \text{MED}_{PFS}$ as the most important and total number of deaths as the second most important diagnostic measure for significant HR_{OS} .

Next, we performed classification tree analysis only on the 58 cases where information were available for both $\% \Delta \text{MED}_{PFS}$ and total number of deaths. The classification tree results are displayed in Figure 2. In 39 cases, increases in median PFS were shorter than 48.27%; and only 3 of these cases showed significant HR_{OS} . There were 19 cases with increases in median PFS reported at least 48.27%. Of these 19 cases, in 12 cases total deaths reported were less than 227 and 5 of them reported significant HR_{OS} . In remaining 7 cases total deaths were 227 or more and 6 of them reported significant HR_{OS} . In Figure 3, comparisons are displayed in terms of $\% \Delta \text{MED}_{PFS}$ and total number of deaths. Figure 3 suggests that the treatment to control comparisons reporting statistically significant HR_{OS} tend to have about 50% or higher median PFS increase and total number deaths 227 or more. The findings of Figure 2 and Figure 3 can be summarized as follows: There is only small chance that study would produce significant HR_{OS} if increase in median PFS is less than 48.27%. Trials with median PFS increase of at least 48.27% seem to have better chance of producing statistically significant HR_{OS} , and having a total of 227 or more OS events further improves the likelihood of obtaining significant HR_{OS} .

Discussion

That a substantial improvement in PFS may be predictive of a corresponding difference in OS makes common sense. However, what is often not obvious is the magnitude of PFS difference that is required to be reasonably confident of observing a statistically significant HR_{OS} . This is crucial in late phase trials where therapeutic agents are being tested and the sponsor needs to decide whether the observed PFS difference could be predictive of a significant and clinically meaningful difference in OS and would merit a marketing authorization application (MAA). The ROC and classification tree analyses employed here are very well suited for such a determination. For example, the ROC approach gives us an overall assessment of diagnostic accuracy based on the AUC metric. On the other hand, the classification tree approach is helpful in identifying non-linear association and influence of other factor, such as total number of deaths, in an interpretable and visible manner. Both the approaches help us to choose an optimal operating point to guide the decision-making process.

Our study findings can be summarized as follows: First of all, ΔMED_{PFS} (i.e. percentage difference in median PFS) is a relatively better diagnostic predictor compared to HR_{PFS} . This is suggested by both ROC analysis and classification tree analyses. Secondly, higher $\% \Delta \text{MED}_{PFS}$ tends to be associated with significant HR_{OS} . However, our classification tree result suggests that higher number of deaths is also important in achieving significant HR_{OS} . The fact that the number of deaths influences the result of statistical testing of HR_{OS} is very logical as deaths are considered as events in OS analysis and increased number of events improves the chance of statistical significance [i.e., power] in survival analyses. Based on our study results, the surrogate threshold effect (STE) in terms of $\% \Delta \text{MED}_{PFS}$ appears to be close to 50% increase in median PFS (ROC analysis: $\text{STE}=44.83\%$; Classification tree analysis: $\text{STE}=48.27\%$). This suggests the studies with 50% improvement in me-

dian PFS are more likely to produce significant HR_{OS} if the total number of OS events at the time analysis is 227 or more. This result can be useful in the context of breast cancer trials in at least two ways: first, a trial showing about 50% increase in median PFS may serve as a useful indicator of statistically significant HR_{OS} while awaiting for OS data to mature. Secondly, if a prospective clinical trial plans to show statistically significant HR_{OS} , then that trial should plan for approximately 227 OS events.

An alternate approach to assessing diagnostic accuracy is the use of Positive predictive value (PPV) and Negative predictive value (NPV) curves introduced by Moskowitz and Pepe [93]. While Positive predictive value (PPV) and negative predictive value (NPV) are great tools, these are not intrinsic measures of diagnostic accuracy, being dependent on the prevalence. In our case, pre-valence refers to prevalence of comparisons with statistically significant HR_{OS} . In fact, Moskowitz and Pepe (2004), when commenting on the applicability of their method in the discussion section of their paper note "Study designs that result in a sample prevalence that does not accurately reflect the true population prevalence will produce biased estimates of PPV and NPV". In their paper, they assume that the data arose from a cohort type design where the sample estimates would be valid. Since we are selecting a sample of studies from the whole population of studies using certain search criteria, there is no reason to assume that the sample prevalence would match the true population prevalence in this case. Also, given the relatively small number of positive studies, we believe that there is not enough data to draw a reliable PPV curve along the lines of Moskowitz and Pepe (2004). While we do acknowledge the intuitive appeal of the PPV and NPV curves, we would like to emphasize that the AUC would be a more appropriate metric for assessing diagnostic accuracy for our data.

There have been numerous studies examining the strength of surrogacy and the ability of a surrogate measure such as PFS to predict OS. Most of the previous meta-analytic studies on metastatic breast cancer attempted to measure the association between improvement in OS with improvement in PFS and clinical benefit rate (CBR) [11, 7, 12, 13, 8, 14]. For example, Miksad *et al.* [12] found only moderate correlation between HR_{PFS} and HR_{OS} (R^2 ranging between 0.35 to 0.59) in taxane and anthracycline based therapies in breast cancer patients. Recently, Amiri-Kordestani *et al.* (2016) [16] reported a moderate association between odds ratio (OR) of CBR and HR_{PFS} ($R^2 = 0.52$), but failed to show any association with OR of CBR with HR_{OS} ($R^2 = 0.01$) from 13 prospective studies submitted to FDA. Our work takes a markedly different path. First, we are interested in assessing the association of treatment effect on PFS with the ultimate interest being the observation or non-observation of a significant HR_{OS} , whereas previous meta-analyses were mostly focused on exploring the association between the treatment effects on PFS and OS. Secondly, we have also considered PFS measures beyond HR_{PFS} such as ΔMED_{PFS} (i.e. difference in median PFS) or $\% \Delta MED_{PFS}$ (i.e., percentage difference in median PFS). Thirdly, we have used fully data dependent non-parametric approaches like empirical ROC curves and classification tree for the meta-analyses which, as far as we know, have not been used before in applications of this type. Fourthly, we have also included other factors such as total sample size and total number of deaths in evaluating the association between PFS measures and significance of HR_{OS} . Last but not the least, we have included a comprehensive list of all published studies since 2000 and hence the number of studies included in this current investigation is relatively higher than previous published meta-analyses.

Our study has its limitations. Not all breast cancer studies published during the period under consideration could be included, although the number is still quite high compared to previous meta

analyses. The selection of studies was driven by common-sense, objective search criteria which were rigorously applied. Thus, only 74 treatment to control comparisons from 65 prospective published trials on breast cancer met the eligibility. The included trials were diverse in nature in terms of patient population considered – some examples of patient population are patients with at least two prior chemotherapies, anthracycline or taxane resistant patients, post-menopausal patients, HER2 positive patients, HER2 negative patients, just to name a few. Further, the included studies were also diverse with regards to tumor type, tumor stage and drug under investigation. We have reported only overall results from non-parametric analyses including all 74 studies; we did not have enough studies to assess surrogacy in each of the categories formed by the combination of these different factors. Thus, the reported overall results may not be applicable in all set-up of breast cancer studies. There is also potential for publication bias. To minimize the publication bias, we have considered all the studies which were published and listed in PubMed database and met our search criteria. Lastly, we have not looked into other cancer indications, although we cannot think of any reason why the method would not be a useful tool in assessing surrogacy of PFS (or any other continuous time-to-event endpoint, for that matter) in other indications and settings.

To conclude, empirical ROC curves and classification trees could be useful tools for assessing how well treatment effect on PFS predicts whether HR_{OS} would be statistically significant or not in breast cancer studies.

References

- [1] Advanced search results from <https://clinicaltrials.gov/ct2/search/advanced>, accessed on 12-Dec-2015.
- [2] Saad, E., Katz, A., Hoff, P., and Buyse, M. (2010). Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. *Annals of oncology*, 21(1):7–12.
- [3] Smith, I. (2006). Goals of treatment for patients with metastatic breast cancer. In *Seminars in oncology*, volume 33, pages 2–5. Elsevier.
- [4] FDA CDER and CBER (May 2007). Clinical trial endpoints for the approval of cancer drugs and biologics.
- [5] Booth, C. M., and Eisenhauer, E. A. (2012). Progression-free survival: meaningful or simply measurable?. *Journal of Clinical Oncology*, 30(10), 1030-1033.
- [6] Venook, A. P. and Tabernero, J. (2014). Progression-free survival: Helpful biomarker or clinically meaningless end point? *Journal of Clinical Oncology*, pages JCO–2014.
- [7] Hackshaw, A., Knight, A., Barrett-Lee, P., & Leonard, R. (2005). Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. *British journal of cancer*, 93(11), 1215-1221.
- [8] Burzykowski, T., Buyse, M., Piccart-Gebhart, M.J., Sledge, G., Carmichael, J., LÅijck, H.J., Mackey, J.R., Nabholz, J.M., Paridaens, R., Biganzoli, L. and Jassem, J. (2008). Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *Journal of Clinical Oncology*, 26(12):1987–1992.

- [9] Michiels, S., Pugliano, L., Marguet, S., Grun, D., Barinoff, J., Cameron, D., Cobleigh, M., Di Leo, A., Johnston, S., Gasparini, G. and Kaufman, B. (2016). Progression-free survival as surrogate end point for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer. *Annals of Oncology*, 27(6): 1029–1034.
- [10] Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*, 8(4):431–440.
- [11] Bruzzi, P., Del Mastro, L., Sormani, M. P., Bastholt, L., Danova, M., Focan, C., and others. (2005). Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *Journal of clinical oncology*, 23(22), 5117–5125.
- [12] Miksad, R. A., Zietemann, V., Gothe, R., Schwarzer, R., Conrads-Frank, A., Schnell-Inderst, P., and others (2008). Progression-free survival as a surrogate endpoint in advanced breast cancer. *International journal of technology assessment in health care*, 24(04), 371–383.
- [13] Sherrill, B., Amonkar, M., Wu, Y., Hirst, C., Stein, S., Walker, M., and Cuzick, J. (2008). Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. *British journal of cancer*, 99(10), 1572–1578.
- [14] Ng, R., Pond, G. R., Tang, P. A., MacIntosh, P. W., Siu, L. L., and Chen, E. X. (2008). Correlation of changes between 2-year disease-free survival and 5-year overall survival in adjuvant breast cancer trials from 1966 to 2006. *Annals of oncology*, 19(3), 481–486.
- [15] Sherrill, B., Kaye, J. A., Sandin, R., Cappelleri, J. C., and Chen, C. (2012). Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther*, 5, 287–296.
- [16] Amiri-Kordestani, L., Cheng, J., Zhang, L., Tang, S., Sridhara, R., Ibrahim, A., ... & Pazdur, R. (2016, May). Association of clinical benefit rate (CBR) with survival: A pooled-analysis of metastatic breast cancer (MBC) trials submitted to the US Food and Drug Administration (FDA). In *ASCO Annual Meeting Proceedings*, Vol. 34, No. 15 suppl, p. e18091).
- [17] Burzykowski, T., and Buyse, M. (2006). Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics*, 5(3): 173–186.
- [18] Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D. and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, 1(1):49–67.
- [19] Buyse, M., Burzykowski, T., Carroll, K., Michiels, S., Sargent, D. J., Miller, L. L., Elfring J. P. and Piedbois, P. (2007). Progression-free survival is a surrogate for survival in advanced colorectal cancer. *Journal of Clinical Oncology*, 25(33): 5218–5224.
- [20] Wang, J., Xu, B., Yuan, P., Ma, F., Li, Q., Zhang, P., Cai, R., Fan, Y., Luo, Y., and Li, Q. (2015). Capecitabine combined with docetaxel versus vinorelbine followed by capecitabine maintenance medication for first-line treatment of patients with advanced breast cancer: Phase 3 randomized trial. *Cancer*, 121(19):3412–3421.
- [21] Pivot, X., Manikhas, A., Åżurawski, B., Chmielowska, E., Karaszewska, B., Allerton, R., Chan, S., Fabi, A., Bidoli, P., Gori, S., and others (2015). CEREBEL (EGF111438): A phase III, randomized, open-label study of lapatinib plus capecitabine versus trastuzumab plus capecitabine

in patients with human epidermal growth factor receptor 2 – positive metastatic breast cancer. *Journal of Clinical Oncology*, 33(14):1564–1573.

- [22] Kaufman, P. A., Awada, A., Twelves, C., Yelle, L., Perez, E. A., Velikova, G., Olivo, M. S., He, Y., Dutcus, C. E., and Cortes, J. (2015). Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. *Journal of clinical oncology*, pages JCO–2013.
- [23] Luck, H.-J., Lubbe, K., Reinisch, M., Maass, N., Feisel-Schwickardi, G., TomÁl, O., Janni, W., Aydogdu, M., NeunhÄuffer, T., Ober, A., and others (2015). Phase III study on efficacy of taxanes plus bevacizumab with or without capecitabine as first-line chemotherapy in metastatic breast cancer. *Breast cancer research and treatment*, 149(1):141–149.
- [24] Martin, M., Loibl, S., von Minckwitz, G., Morales, S., Martinez, N., Guerrero, A., Anton, A., Aktas, B., Schoenegg, W., MuÁsoz, M., and others (2015). Phase III Trial Evaluating the Addition of Bevacizumab to Endocrine Therapy As First-Line Treatment for Advanced Breast Cancer: The Letrozole/Fulvestrant and Avastin (LEA) Study. *Journal of Clinical Oncology*, pages JCO–2014.
- [25] Krop, I. E., Lin, N. U., Blackwell, K., Guardino, E., Huober, J., Lu, M., Miles, D., Samant, M., Welslau, M., and Dieras, V. (2015). Trastuzumab emtansine (T-DM1) versus lapatinib plus capecitabine in patients with HER2-positive metastatic breast cancer and central nervous system metastases: a retrospective, exploratory analysis in EMILIA. *Annals of Oncology*, 26(1):113–119.
- [26] Dieras, V., Campone, M., Yardley, D. A., Romieu, G., Valero, V., Isakoff, S. J., Koeppen, H., Wilson, T. R., Xiao, Y., Shames, D. S., and others (2015). Randomized, phase II, placebo-controlled trial of onartuzumab and/or bevacizumab in combination with weekly paclitaxel in patients with metastatic triple-negative breast cancer. *Annals of Oncology*, 26(9):1904–1910.
- [27] Rugo, H. S., Barry, W. T., Moreno-Aspitia, A., Lyss, A. P., Cirrincione, C., Leung, E., Mayer, E. L., Naughton, M., Toppmeyer, D., Carey, L. A., and others (2015). Randomized phase III trial of paclitaxel once per week compared with nanoparticle albumin-bound nab-paclitaxel once per week or ixabepilone with bevacizumab as first-line chemotherapy for locally recurrent or metastatic breast cancer: CALGB 40502/NCCTG N063h (Alliance). *Journal of Clinical Oncology*, 33(21):2361–2369.
- [28] Clemens, M. R., Gladkov, O. A., Gartner, E., Vladimirov, V., Crown, J., Steinberg, J., Jie, F., and Keating, A. (2015). Phase II, multicenter, open-label, randomized study of YM155 plus docetaxel as first-line treatment in patients with HER2-negative metastatic breast cancer. *Breast cancer research and treatment*, 149(1):171–179.
- [29] Mackey, J. R., Ramos-Vazquez, M., Lipatov, O., McCarthy, N., Krasnozhon, D., Semiglazov, V., Manikhas, A., Gelmon, K. A., Konecny, G. E., Webster, M., and others (2015). Primary results of ROSE/TRIO-12, a randomized placebo-controlled phase III trial evaluating the addition of ramucirumab to first-line docetaxel chemotherapy in metastatic breast cancer. *Journal of Clinical Oncology*, pages JCO–2014.
- [30] Janni, W., Sarosiek, T., Karaszewska, B., Pikiel, J., Staroslawska, E., Potemski, P., Salat, C., Brain, E., Caglevic, C., Briggs, K., and others (2014). A phase II, randomized, multicenter study evaluating the combination of lapatinib and vinorelbine in women with ErbB2 overexpressing metastatic breast cancer. *Breast cancer research and treatment*, 143(3):493–505.

- [31] Smorenburg, C. H., de Groot, S. M., van Leeuwen-Stok, A. E., Hamaker, M. E., Wymenga, A. N., de Graaf, H., de Jongh, F. E., Braun, J. J., Los, M., Maartense, E., and others (2014). A randomized phase III study comparing pegylated liposomal doxorubicin with capecitabine as first-line chemotherapy in elderly patients with metastatic breast cancer: results of the OMEGA study of the Dutch Breast Cancer Research Group BOOG. *Annals of oncology*, page mdt588.
- [32] Burstein, H. J., Cirincione, C. T., Barry, W. T., Chew, H. K., Tolane, S. M., Lake, D. E., Ma, C., Blackwell, K. L., Winer, E. P., and Hudis, C. A. (2014). Endocrine Therapy With or Without Inhibition of Epidermal Growth Factor Receptor and Human Epidermal Growth Factor Receptor 2: A Randomized, Double-Blind, Placebo-Controlled Phase III Trial of Fulvestrant With or Without Lapatinib for Postmenopausal Women With Hormone Receptor 2 – Positive Advanced Breast Cancer – CALGB 40302 (Alliance). *Journal of Clinical Oncology*, 32(35):3959–3966.
- [33] Kim, S.-B., Yoo, C., Ro, J., Im, S.-A., Im, Y.-H., Kim, J. H., Ahn, J.-H., Jung, K. H., Song, H. S., Kang, S. Y., and others (2014). Combination of docetaxel and TSU-68, an oral antiangiogenic agent, in patients with metastatic breast cancer previously treated with anthracycline: Randomized phase II multicenter trial. *Investigational new drugs*, 32(4):753–761.
- [34] Piccart, M., Hortobagyi, G. N., Campone, M., Pritchard, K. I., Lebrun, F., Ito, Y., Noguchi, S., Perez, A., Rugo, H. S., Deleu, I., and others (2014). Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2. *Annals of Oncology*, page mdu456.
- [35] O’Shaughnessy, J., Schwartzberg, L., Danso, M. A., Miller, K. D., Rugo, H. S., Neubauer, M., Robert, N., Hellerstedt, B., Saleh, M., Richards, P., and others (2014). Phase III study of iniparib plus gemcitabine and carboplatin versus gemcitabine and carboplatin in patients with metastatic triple-negative breast cancer. *Journal of Clinical Oncology*, pages JCO–2014.
- [36] Baselga, J., Manikhas, A., Cortés, J., Llombart, A., Roman, L., Semiglazov, V. F., Byakhov, M., Lokanatha, D., Forenza, S., Goldfarb, R. H., and others (2014). Phase III trial of nonpegylated liposomal doxorubicin in combination with trastuzumab and paclitaxel in HER2-positive metastatic breast cancer. *Annals of oncology*, 25(3):592–598.
- [37] Crown, J. P., Dieras, V., Staroslawska, E., Yardley, D. A., Bachelot, T., Davidson, N., Wildiers, H., Fasching, P. A., Capitain, O., Ramos, M., and others (2013). Phase III trial of sunitinib in combination with capecitabine versus capecitabine monotherapy for the treatment of patients with pretreated metastatic breast cancer. *Journal of Clinical Oncology*, pages JCO–2012.
- [38] Hurvitz, S. A., Dirix, L., Kocsis, J., Bianchi, G. V., Lu, J., Vinholes, J., Guardino, E., Song, C., Tong, B., Ng, V., and others (2013). Phase II randomized study of trastuzumab emtansine versus trastuzumab plus docetaxel in patients with human epidermal growth factor receptor 2 – positive metastatic breast cancer. *Journal of Clinical Oncology*, pages JCO–2012.
- [39] Blumenthal, G. M., Scher, N. S., Cortazar, P., Chattopadhyay, S., Tang, S., Song, P., Liu, Q., Ringgold, K., Pilaro, A. M., Tilley, A., and others (2013). First FDA approval of dual anti-HER2 regimen: pertuzumab in combination with trastuzumab and docetaxel for HER2-positive metastatic breast cancer. *Clinical Cancer Research*, 19(18):4911–4916.
- [40] Park, Y. H., Jung, K. H., Im, S. A., Sohn, J. H., Ro, J., Ahn, J. H., and others (2013). Phase III, multicenter, randomized trial of maintenance chemotherapy versus observation in patients

with metastatic breast cancer after achieving disease control with six cycles of gemcitabine plus paclitaxel as first-line chemotherapy: KCSG-BR07-02. *Journal of Clinical Oncology*, JCO-2012.

- [41] Luck, H.-J., Du Bois, A., Loibl, S., Schrader, I., Huober, J., Heilmann, V., Beckmann, M., StÄdhler, A., Jackisch, C., Hubalek, M., and others (2013). Capecitabine plus paclitaxel versus epirubicin plus paclitaxel as first-line treatment for metastatic breast cancer: efficacy and safety results of a randomized, phase III trial by the AGO Breast Cancer Study Group. *Breast cancer research and treatment*, 139(3):779–787.
- [42] Kader, Y. A., Spielmann, M., El-Nahas, T., Sakr, A., and Metwally, H. (2013). Comparative study analyzing survival and safety of bevacizumab/carboplatin/paclitaxel versus carboplatin/docetaxel in initial treatment of metastatic Her-2-negative breast cancer. *Breast Cancer: Targets and Therapy*, 5:37.
- [43] Harvey, V. J., Sharples, K. J., Isaacs, R. J., Jameson, M. B., Jeffery, G. M., McLaren, B. R., Pollard, S., Riley, G. A., Simpson, A. B., Hinder, V. A., and others (2013). A randomized phase II study comparing capecitabine alone with capecitabine and oral cyclophosphamide in patients with advanced breast cancer-cyclo II. *Annals of oncology*, page mdt065.
- [44] Gianni, L., Romieu, G. H., Lichinitser, M., Serrano, S. V., Mansutti, M., Pivot, X., Mariani, P., Andre, F., Chan, A., Lipatov, O., and others (2013). AVEREL: a randomized phase III Trial evaluating bevacizumab in combination with docetaxel and trastuzumab as first-line therapy for HER2- positive locally recurrent/metastatic breast cancer. *Journal of Clinical Oncology*, 31(14):1719–1725.
- [45] Guan, Z., Xu, B., DeSilvio, M. L., Shen, Z., Arpornwirat, W., Tong, Z., Lorvidhaya, V., Jiang, Z., Yang, J., Makhson, A., and others (2013). Randomized trial of lapatinib versus placebo added to paclitaxel in the treatment of human epidermal growth factor receptor 2 – overexpressing metastatic breast cancer. *Journal of Clinical Oncology*, 31(16):1947–1953.
- [46] Lee, C. K., GebSKI, V. J., Coates, A. S., Veillard, A.-S., Harvey, V., Tattersall, M. H., Byrne, M. J., Brigham, B., Forbes, J., Simes, R. J., and others (2013). Trade-offs in quality of life and survival with chemotherapy for advanced breast cancer: mature results of a randomized trial comparing single-agent mitoxantrone with combination cyclophosphamide, methotrexate, 5-fluorouracil and prednisone. *SpringerPlus*, 2(1):1–10.
- [47] Yamamoto, Y., Ishikawa, T., Hozumi, Y., Ikeda, M., Iwata, H., Yamashita, H., Toyama, T., Chishima, T., Saji, S., Yamamoto-Ibusuki, M., and others (2013). Randomized controlled trial of toremifene 120 mg compared with exemestane 25 mg after prior treatment with a non-steroidal aromatase inhibitor in postmenopausal women with hormone receptor-positive metastatic breast cancer. *BMC cancer*, 13(1):239.
- [48] Yardley, D. A., Ismail-Khan, R. R., Melichar, B., Lichinitser, M., Munster, P. N., Klein, P. M., Cruickshank, S., Miller, K. D., Lee, M. J., and Trepel, J. B. (2013). Randomized phase II, double-blind, placebo-controlled study of exemestane with or without entinostat in postmenopausal women with locally recurrent or metastatic estrogen receptor-positive breast cancer progressing on treatment with a nonsteroidal aromatase inhibitor. *Journal of Clinical Oncology*, 31(17):2128–2135.

- [49] Curigliano, G., Pivot, X., Cortes, J., Elias, A., Cesari, R., Khosravan, R., Collier, M., Huang, X., Cataruozolo, P. E., Kern, K. A., and others (2013). Randomized phase II study of sunitinib versus standard of care for patients with previously treated advanced triple negative breast cancer. *The Breast*, 22(5):650–656.
- [50] Baselga, J., Gomez, P., Greil, R., Braga, S., Climent, M. A., Wardley, A. M., Kaufman, B., Stemmer, S. M., Pego, A., Chan, A., and others (2013). Randomized phase II study of the anti-epidermal growth factor receptor monoclonal antibody cetuximab with cisplatin versus cisplatin alone in patients with metastatic triple-negative breast cancer. *Journal of clinical oncology*, 31(20):2586–2592.
- [51] Fan, Y., Xu, B. H., Yuan, P., Ma, F., Wang, J. Y., Ding, X. Y., Zhang, P., Li, Q., and Cai, R. G. (2013). Docetaxel–cisplatin might be superior to docetaxel–capecitabine in the first-line treatment of metastatic triple-negative breast cancer. *Annals of oncology*, page mds603.
- [52] Baselga, J., Segalla, J. G. M., Roche, H., del Giglio, A., Pinczowski, H., Ciruelos, E. M., Cabral Filho, S., Gomez, P., Van Eyll, B., Bermejo, B., and others (2012). Sorafenib in combination with capecitabine: an oral regimen for patients with HER2-negative locally advanced or metastatic breast cancer. *Journal of clinical oncology*, 30(13):1484–1491.
- [53] Pallis, A., Boukovinas, I., Ardavanis, A., Varthalitis, I., Malamos, N., Georgoulas, V., and Mavroudis, D. (2012). A multicenter randomized phase iii trial of vinorelbine/gemcitabine doublet versus capecitabine monotherapy in anthracycline- and taxane-pretreated women with metastatic breast cancer. *Annals of oncology*, 23(5):1164–1169.
- [54] Donoghue, M., Lemery, S. J., Yuan, W., He, K., Sridhara, R., Shord, S., Zhao, H., Marathe, A., Kotch, L., Jee, J., and others (2012). Eribulin mesylate for the treatment of patients with refractory metastatic breast cancer: Use of a “physician’s choice” control arm in a randomized approval trial. *Clinical Cancer Research*, 18(6):1496–1505.
- [55] Bergh, J., Bondarenko, I. M., Lichinitser, M. R., Liljegren, A., Greil, R., Voytko, N. L., Makhson, A. N., Cortes, J., Lortholary, A., Bischoff, J., and others (2012). First-line treatment of advanced breast cancer with sunitinib in combination with docetaxel versus docetaxel alone: results of a prospective, randomized phase III study. *Journal of Clinical Oncology*, pages JCO–2011.
- [56] Hatschek, T., Carlsson, L., Einbeigi, Z., Lidbrink, E., Linderholm, B., Lindh, B., Loman, N., Malmberg, M., Rotstein, S., Soderberg, M., and others (2012). Individually tailored treatment with epirubicin and paclitaxel with or without capecitabine as first-line chemotherapy in metastatic breast cancer: a randomized multicenter trial. *Breast cancer research and treatment*, 131(3):939–947.
- [57] Roche, H., Conte, P., Perez, E. A., Sparano, J. A., Xu, B., Jassem, J., Peck, R., Kelleher, T., and Hortobagyi, G. N. (2011). Ixabepilone plus capecitabine in metastatic breast cancer patients with reduced performance status previously treated with anthracyclines and taxanes: a pooled analysis by performance status of efficacy and safety data from 2 phase III studies. *Breast cancer research and treatment*, 125(3):755–765.
- [58] Brufsky, A., Hoelzer, K., Beck, T., Whorf, R., Keaton, M., Nadella, P., Krill-Jackson, E., Kroener, J., Middleman, E., Frontiera, M., and others (2011a). A randomized phase II study of paclitaxel and bevacizumab with and without gemcitabine as first-line treatment for metastatic breast cancer. *Clinical breast cancer*, 11(4):211–220.

- [59] Xu, B., Jiang, Z., Kim, S.-B., Yu, S., Feng, J., Malzyner, A., del Giglio, A., Chung, H. C., Shen, L. J., and Pen, D. L. K. (2011). Biweekly gemcitabine–paclitaxel, gemcitabine–carboplatin, or gemcitabine–cisplatin as first-line treatment in metastatic breast cancer after anthracycline failure: a phase II randomized selection trial. *Breast Cancer*, 18(3):203–212.
- [60] Robert, N. J., Saleh, M. N., Paul, D., Generali, D., Gressot, L., Copur, M. S., Brufsky, A. M., Minton, S. E., Giguere, J. K., Smith, J. W., and others (2011b). Sunitinib plus paclitaxel versus bevacizumab plus paclitaxel for first-line treatment of patients with advanced breast cancer: a phase III, randomized, open-label trial. *Clinical breast cancer*, 11(2):82–92.
- [61] Robert, N. J., Dieras, V., Glaspy, J., Brufsky, A. M., Bondarenko, I., Lipatov, O. N., Perez, E. A., Yardley, D. A., Chan, S. Y., Zhou, X., and others (2011a). RIBBON-1: Randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab for first-line treatment of human epidermal growth factor receptor 2–negative, locally recurrent or metastatic breast cancer. *Journal of Clinical Oncology*, 29(10):1252–1260.
- [62] Brufsky, A. M., Hurvitz, S., Perez, E., Swamy, R., Valero, V., O’Neill, V., and Rugo, H. S. (2011b). RIBBON-2: A randomized, double-blind, placebo-controlled, phase III trial evaluating the efficacy and safety of bevacizumab in combination with chemotherapy for second-line treatment of human epidermal growth factor receptor 2–negative metastatic breast cancer. *Journal of Clinical Oncology*, pages JCO–2010.
- [63] Hamberg, P., Bos, M. M., Braun, H. J., Stouthard, J. M., van Deijk, G. A., Erdkamp, F. L., van der Stelt-Frissen, I. N., Bontenbal, M., Creemers, G.-J. M., Portielje, J. E., and others (2011). Randomized phase II study comparing efficacy and safety of combination-therapy trastuzumab and docetaxel vs. sequential therapy of trastuzumab followed by docetaxel alone at progression as first-line chemotherapy in patients with HER2+ metastatic breast cancer: HERTAX trial. *Clinical breast cancer*, 11(2):103–113.
- [64] Inoue, K., Nakagami, K., Mizutani, M., Hozumi, Y., Fujiwara, Y., Masuda, N., Tsukamoto, F., Saito, M., Miura, S., Eguchi, K., and others (2010). Randomized phase III trial of trastuzumab monotherapy followed by trastuzumab plus docetaxel versus trastuzumab plus docetaxel as first-line therapy in patients with HER2-positive metastatic breast cancer: the JO17360 Trial Group. *Breast cancer research and treatment*, 119(1):127–136.
- [65] Sparano, J. A., Vrdoljak, E., Rixe, O., Xu, B., Manikhas, A., Medina, C., Da Costa, S. C. V., Ro, J., Rubio, G., Rondinon, M., and others (2010). Randomized phase III trial of ixabepilone plus capecitabine versus capecitabine in patients with metastatic breast cancer previously treated with an anthracycline and a taxane. *Journal of clinical oncology*, pages JCO–2009.
- [66] Barrios, C. H., Liu, M.-C., Lee, S. C., Vanlemmens, L., Ferrero, J.-M., Tabei, T., Pivot, X., Iwata, H., Aogi, K., Lugo-Quintana, R., and others (2010). Phase III randomized trial of sunitinib versus capecitabine in patients with previously treated HER2-negative advanced breast cancer. *Breast cancer research and treatment*, 121(1):121–131.
- [67] Blackwell, K. L., Burstein, H. J., Storniolo, A. M., Rugo, H., Sledge, G., Koehler, M., Ellis, C., Casey, M., Vukelja, S., Bischoff, J., and others (2010). Randomized study of Lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. *Journal of Clinical Oncology*, 28(7):1124–1130.

- [68] Moulder, S. L., Holmes, F. A., Tolcher, A. W., Thall, P., Broglio, K., Valero, V., Buzdar, A. U., Arbuck, S. G., Seidman, A., and Hortobagyi, G. N. (2010). A randomized phase 2 trial comparing 3-hour versus 96-hour infusion schedules of paclitaxel for the treatment of metastatic breast cancer. *Cancer*, 116(4):814–821.
- [69] Clemons, M., Joy, A. A., Abdalnabi, R., Kotliar, M., Lynch, J., Jordaan, J. P., Iscoe, N., and Gelmon, K. (2010). Phase II, double-blind, randomized trial of capecitabine plus enzastaurin versus capecitabine plus placebo in patients with metastatic or recurrent breast cancer after prior anthracycline and taxane therapy. *Breast cancer research and treatment*, 124(1):177–186.
- [70] Ruff, P., Vorobiof, D. A., Jordaan, J. P., Demetriou, G. S., Moodley, S. D., Nosworthy, A. L., Werner, I. D., Raats, J., and Burgess, L. J. (2009). A randomized, placebo-controlled, double-blind phase 2 study of docetaxel compared to docetaxel plus zosuquidar (LY335979) in women with metastatic or locally recurrent breast cancer who have received one prior chemotherapy regimen. *Cancer chemotherapy and pharmacology*, 64(4):763–768.
- [71] Kaufman, B., Mackey, J. R., Clemens, M. R., Bapsy, P. P., Vaid, A., Wardley, A., Tjulandin, S., Jahn, M., Lehle, M., Feyereislova, A., and others (2009). Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2–positive, hormone receptor–positive metastatic breast cancer: Results from the randomized phase III TAnDEM study. *Journal of Clinical Oncology*, 27(33):5529–5537.
- [72] Cassier, P. A., Chabaud, S., Trillet-Lenoir, V., Peaud, P.-Y., Tigaud, J.-D., Cure, H., Orfeuvre, H., Salles, B., Martin, C., Jacquin, J.-P., and others (2008). A phase-III trial of doxorubicin and docetaxel versus doxorubicin and paclitaxel in metastatic breast cancer: results of the ERASME 3 study. *Breast cancer research and treatment*, 109(2): 343–50.
- [73] Aapro, M., Leonard, R. C., Barnadas, A., Marangolo, M., Untch, M., Malamos, N., Mayordomo, J., Reichert, D., Pedrini, J. L., Ukarma, L., and others (2008). Effect of once-weekly epoetin beta on survival in patients with metastatic breast cancer receiving anthracycline-and/or taxane-based chemotherapy: results of the breast cancer anemia and the value of erythropoietin (brave) study. *Journal of Clinical Oncology*, 26(4):592–598.
- [74] Cameron, D., Casey, M., Press, M., Lindquist, D., Pienkowski, T., Romieu, C., Chan, S., Jagiello-Gruszfeld, A., Kaufman, B., Crown, J. and others (2008). A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. *Breast cancer research and treatment*, 112(3): 533–543.
- [75] Rivera, E., Mejia, J. A., Arun, B. K., Adinin, R. B., Walters, R. S., Brewster, A., Broglio, K. R., Yin, G., Esmali, B., Hortobagyi, G. N., and others (2008). Phase 3 study comparing the use of docetaxel on an every-3-week versus weekly schedule in the treatment of metastatic breast cancer. *Cancer*, 112(7):1455–1461.
- [76] Crump, M., Gluck, S., Tu, D., Stewart, D., Levine, M., Kirkbride, P., Dancey, J., O’Reilly, S., Shore, T., Couban, S., and others (2008). Randomized trial of high-dose chemotherapy with autologous peripheral-blood stem-cell support compared with standard-dose chemotherapy in women with metastatic breast cancer: NCIC MA. 16. *Journal of Clinical Oncology*, 26(1):37–43.

- [77] Miller, K., Wang, M., Gralow, J., Dickler, M., Cobleigh, M., Perez, E. A., Shenkier, T., Cella, D., and Davidson, N. E. (2007). Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *New England Journal of Medicine*, 357(26):2666–2676.
- [78] Kroger, N., Frick, M., Gluz, O., Mohrmann, S., Metzner, B., Jackisch, C., Ko, Y., Lindemann, H.-W., Meier, C. R., Lohrmann, H. P., and others (2006). Randomized trial of single compared with tandem high-dose chemotherapy followed by autologous stem-cell transplantation in patients with chemotherapy-sensitive metastatic breast cancer. *Journal of clinical oncology*, 24(24):3919–3926.
- [79] Miller, K. D., Chap, L. I., Holmes, F. A., Cobleigh, M. A., Marcom, P. K., Fehrenbacher, L., Dickler, M., Overmoyer, B. A., Reimann, J. D., Sing, A. P., and others (2005). Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. *Journal of Clinical Oncology*, 23(4):792–799.
- [80] Sparano, J. A., Bernardo, P., Stephenson, P., Gradishar, W. J., Ingle, J. N., Zucker, S., and Davidson, N. E. (2004). Randomized phase III trial of marimastat versus placebo in patients with metastatic breast cancer who have responding or stable disease after first-line chemotherapy: Eastern Cooperative Oncology Group trial E2196. *Journal of Clinical Oncology*, 22(23):4683–4690.
- [81] Keller, A. M., Mennel, R. G., Georgoulas, V. A., Nabholz, J.-M., Erazo, A., Lluch, A., Vogel, C. L., Kaufmann, M., von Minckwitz, G., Henderson, I. C., and others (2004). Randomized phase III trial of pegylated liposomal doxorubicin versus vinorelbine or mitomycin C plus vinblastine in women with taxane-refractory advanced breast cancer. *Journal of Clinical Oncology*, 22(19):3893–3901.
- [82] O’Brien, M. E. R., Wigler, N., Inbar, M., Rosso, R., Grischke, E., Santoro, A., Catane, R., Kieback, D. G., Tomczak, P., Ackland, S. P., and others (2004). Reduced cardiotoxicity and comparable efficacy in a phase III trial of pegylated liposomal doxorubicin HCl (CAELYX/Doxil) versus conventional doxorubicin for first-line treatment of metastatic breast cancer. *Annals of oncology*, 15(3):440–449.
- [83] van der Hage, J. A., van de Velde, C. J., Julien, J.-P., Tubiana-Hulin, M., Vandervelden, C., Duchateau, L., and Investigators, C. (2001). Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. *Journal of Clinical Oncology*, 19(22):4224–4237.
- [84] Klijn, J. G., Beex, L. V., Mauriac, L., van Zijl, J. A., Veyret, C., Wildiers, J., Jassem, J., Piccart, M., Burghouts, J., Becquart, D., and others (2000). Combined treatment with buserelin and tamoxifen in premenopausal metastatic breast cancer: a randomized study. *Journal of the National Cancer Institute*, 92(11):903–911.
- [85] Zhou, X.-H., McClish, D. K., and Obuchowski, N. A. (2009). *Statistical methods in diagnostic medicine*, volume 569. John Wiley & Sons.
- [86] Moskowitz, C. S., and Pepe, M. S. (2004). Quantifying and comparing the accuracy of binary biomarkers when predicting a failure time outcome. *Statistics in medicine*, 23(10), 1555–1570.
- [87] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

- [88] Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1):32–35.
- [89] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- [90] Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCr: visualizing classifier performance in R. *Bioinformatics*, 21(20):7881.
- [91] Therneau, T., Atkinson, B., and Ripley, B. (2015). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-9.
- [92] Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22.
- [93] Moskowitz, C. S., and Pepe, M. S. (2004). Quantifying and comparing the accuracy of binary biomarkers when predicting a failure time outcome. *Statistics in medicine*, 23(10), 1555–1570.